

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
4 October 2001 (04.10.2001)

PCT

(10) International Publication Number  
**WO 01/72838 A2**

- (51) International Patent Classification<sup>7</sup>: **C07K 14/705** (74) Agent: **MILLMAN, Robert, A.**; Celera Genomics Corp., 45 West Gude Drive C2-4, Rockville, MD 20850 (US).
- (21) International Application Number: **PCT/US01/09523**
- (22) International Filing Date: **27 March 2001 (27.03.2001)** (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (25) Filing Language: **English**
- (26) Publication Language: **English**
- (30) Priority Data:  
60/192,853 29 March 2000 (29.03.2000) US  
09/635,593 9 August 2000 (09.08.2000) US
- (71) Applicant: **PE CORPORATION (NY)** [US/US]; Robert A. Millman, 761 Main Avenue, Norwalk, CT 06859 (US).
- (72) Inventors: **WEI, Ming-Hui**; c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US). **CRAVCHIK, Anibal**; c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US). **DI FRANCESCO, Valentina**; c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US). **BEASLEY, Ellen, M.**; c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US).
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: ISOLATED HUMAN G-PROTEIN COUPLED RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR PROTEINS, AND USES THEREOF

```
1 ATGGATCCAA CCACCCCGG CTGGGGAACA GAAAGTACAA CAGTGAATGG
51 AAATGACCAA GCCCTTCCTC TGCTTTGTGG CAAGGAGACC CTGATCCCGG
101 TCTTCTCTGAT CCTTTTCATT GCCCTGGTGG GGCTGGTAGG AAACGGGTTT
151 GTGCTCTGGC TCTGCGGCTT CCGCATGCGC AGGAACGCGT TCTCTGTCTA
201 CGTCTCTCAGC CTGGCCGGGG CCGACTTCCT CTTCCTCTGC TTCCAGATTA
251 TAAATTGCCT GGTGTACCTC AGTAACTTCT TCTGTTCCAT CTCCATCAAT
301 TTCCCTAGCT TCTTACCAC TGTGATGACC TGTGCTTACC TTGCAGGCGT
351 GAGCATGCTG AGCACCCTCA GCACCGAGCG CTGCTGTGCC GTCTGTGGC
401 CCATCTGGTA TCGCTGCCGC CGCCCGAGAC ACCTGTGCGC GGTGCTGTGT
451 GTCTGTGCTC GGGCCCTGTC CTTACTGTGT AGCATCTTGG AAGGGAAGTT
501 CTGTGGCTTC TTATTTAGTG ATGCTGACTC TGGTTGGTGT CAGACATTTG
551 ATTTTCATCAC TGCAGCGTGG CTGAITTTT TATTCTGGT TCTCTGTGGG
601 TCCAGTCTGG CCTGTCTGGT CAGGATCCTC TGTGGCTCCA GGGGCTCTGC
651 ACTGACGAGG CTGTACCTGA CCATCTGCTC CACAGTGTGT GTGTTCTCTC
701 TCTGCGGCGT GCCCTTGGC ATTCACTGGT TCCTAATATT ATGGATCTGG
751 AAGGATCTGT ATGCTTATT TTGTCAATTT CATCCAGTTT CAGTTGTCTC
801 GTCATCTCTT AACAGCAGTG CCAACCCCAT CATTTACTTC TTCTGTGGCT
851 CTTTTAGGAA GCAGTGGGGG CTGCAGCAGC CGATCTCTCA GCTGCTCTC
901 CAGAGGGGCT TGCAGGACAT TGCTGAGGTG GATCAGAGTG AAGGATGCTT
951 CCGTCAGGGC ACCCCGGAGA TGTCGAGAAG CAGTCTGGTG TAG (SEQ ID NO:1)
```

FEATURES:  
Start: 1  
Stop: 991

HOMOLOGOUS PROTEIN:

Top BLAST hits:  
g11134079|sp|P23749|RTA RAT PROBABLE G PROTEIN-COUPLED RECEPTOR... 125 7e-28  
g116981186|ref|NP\_036895.1|| MAS1 oncogene >g1135921|sp|P12526... 111 1e-23  
g1145051051|ref|NP\_002368.1|| MAS1 oncogene >g1135920|sp|P04201... 109 6e-23  
g116678804|ref|NP\_032578.1|| MAS1 oncogene >g1266505|sp|P30554... 106 5e-22

Tissue expression for modulatory use (from cDNA screening panel):

Human Uterus  
Human Testis  
Human Heart  
Human mammary gland  
Human liver

(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the Human genome, the GPCR peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the GPCR peptides and methods of identifying modulators of the GPCR peptides.

WO 01/72838 A2



*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## ISOLATED HUMAN G-PROTEIN COUPLED RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR PROTEINS, AND USES THEREOF

### RELATED APPLICATIONS

5       The present application claims priority to U.S. Serial Nos. 60/192,853 (Atty. Docket CL000410-PROV), filed March 29, 2000, and 09/635,593, filed August 9, 2000 (Atty. Docket CL000751).

### FIELD OF THE INVENTION

10       The present invention is in the field of G-Protein coupled receptors (GPCRs) that are related to the MAS proto-oncogene receptor subfamily, recombinant DNA molecules and protein production. The present invention specifically provides novel GPCR peptides and proteins and nucleic acid molecules encoding such protein molecules, for use in the development of human therapeutics and human therapeutic development, particularly to control tumors that express the  
15    GPCR.

### BACKGROUND OF THE INVENTION

#### G-protein coupled receptors

G-protein coupled receptors (GPCRs) constitute a major class of proteins responsible for  
20    transducing a signal within a cell. GPCRs have three structural domains: an amino terminal extracellular domain, a transmembrane domain containing seven transmembrane segments, three extracellular loops, and three intracellular loops, and a carboxy terminal intracellular domain. Upon binding of a ligand to an extracellular portion of a GPCR, a signal is transduced within the cell that results in a change in a biological or physiological property of the cell. GPCRs, along with G-proteins  
25    and effectors (intracellular enzymes and channels modulated by G-proteins), are the components of a modular signaling system that connects the state of intracellular second messengers to extracellular inputs.

GPCR genes and gene-products are potential causative agents of disease (Spiegel *et al.*, *J. Clin. Invest.* 92:1119-1125 (1993); McKusick *et al.*, *J. Med. Genet.* 30:1-26 (1993)). Specific defects in the  
30    rhodopsin gene and the V2 vasopressin receptor gene have been shown to cause various forms of retinitis pigmentosum (Nathans *et al.*, *Annu. Rev. Genet.* 26:403-424(1992)), and nephrogenic diabetes

insipidus (Holtzman *et al.*, *Hum. Mol. Genet.* 2:1201-1204 (1993)). These receptors are of critical importance to both the central nervous system and peripheral physiological processes. Evolutionary analyses suggest that the ancestor of these proteins originally developed in concert with complex body plans and nervous systems.

5       The GPCR protein superfamily can be divided into five families: Family I, receptors typified by rhodopsin and the  $\beta$ 2-purinergic receptor and currently represented by over 200 unique members (Dohlman *et al.*, *Annu. Rev. Biochem.* 60:653-688 (1991)); Family II, the parathyroid hormone/calcitonin/secretin receptor family (Juppner *et al.*, *Science* 254:1024-1026 (1991); Lin *et al.*, *Science* 254:1022-1024 (1991)); Family III, the metabotropic glutamate receptor family (Nakanishi, 10 *Science* 258 597:603 (1992)); Family IV, the cAMP receptor family, important in the chemotaxis and development of *D. discoideum* (Klein *et al.*, *Science* 241:1467-1472 (1988)); and Family V, the fungal mating pheromone receptors such as STE2 (Kurjan, *Annu. Rev. Biochem.* 61:1097-1129 (1992)).

There are also a small number of other proteins that present seven putative hydrophobic segments and appear to be unrelated to GPCRs; they have not been shown to couple to G-proteins.

15 *Drosophila* expresses a photoreceptor-specific protein, bride of sevenless (boss), a seven-transmembrane-segment protein that has been extensively studied and does not show evidence of being a GPCR (Hart *et al.*, *Proc. Natl. Acad. Sci. USA* 90:5047-5051 (1993)). The gene *frizzled* (*fz*) in *Drosophila* is also thought to be a protein with seven transmembrane segments. Like boss, *fz* has not been shown to couple to G-proteins (Vinson *et al.*, *Nature* 338:263-264 (1989)).

20       G proteins represent a family of heterotrimeric proteins composed of  $\alpha$ ,  $\beta$  and  $\gamma$  subunits, that bind guanine nucleotides. These proteins are usually linked to cell surface receptors, e.g., receptors containing seven transmembrane segments. Following ligand binding to the GPCR, a conformational change is transmitted to the G protein, which causes the  $\alpha$ -subunit to exchange a bound GDP molecule for a GTP molecule and to dissociate from the  $\beta\gamma$ -subunits. The GTP-bound form of the  $\alpha$ -subunit 25 typically functions as an effector-modulating moiety, leading to the production of second messengers, such as cAMP (e.g., by activation of adenylyl cyclase), diacylglycerol or inositol phosphates. Greater than 20 different types of  $\alpha$ -subunits are known in humans. These subunits associate with a smaller pool of  $\beta$  and  $\gamma$  subunits. Examples of mammalian G proteins include Gi, Go, Gq, Gs and Gt. G proteins are described extensively in Lodish *et al.*, *Molecular Cell Biology*, (Scientific American 30 Books Inc., New York, N.Y., 1995), the contents of which are incorporated herein by reference. GPCRs, G proteins and G protein-linked effector and second messenger systems have been reviewed in *The G-Protein Linked Receptor Fact Book*, Watson *et al.*, eds., Academic Press (1994).

GPCRs, particularly members of the MAS proto-oncogene receptor subfamily, are a major target for drug action and development, particularly to control cancers of cell that express this receptor.

Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown GPCRs. The present invention advances the state of the art by providing a previously unidentified human GPCR.

## 5 SUMMARY OF THE INVENTION

The present invention is based in part on the identification of nucleic acid sequences that encode amino acid sequences of human GPCR peptides and proteins that are related to the MAS proto-oncogene subfamily, allelic variants thereof and other mammalian orthologs thereof. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as  
10 models for the development of human therapeutic targets, aid in the identification of therapeutic proteins and serve as targets for the development of human therapeutic agents.

The proteins of the present inventions are GPCRs that participate in signaling pathways mediated by the MAS proto-oncogene subfamily in cells that express these proteins (see expression information in Figure 1, the GPCR of the present invention is expressed in the uterus, testis, heart, liver  
15 and mammary gland as confirmed by cDNA panel screening). As used herein, a "signaling pathway" refers to the modulation (e.g., stimulation or inhibition) of a cellular function/activity upon the binding of a ligand to the GPCR protein. Examples of such functions include mobilization of intracellular molecules that participate in a signal transduction pathway, e.g., phosphatidylinositol 4,5-bisphosphate (PIP<sub>2</sub>), inositol 1,4,5-triphosphate (IP<sub>3</sub>) and adenylate cyclase; polarization of the plasma membrane;  
20 production or secretion of molecules; alteration in the structure of a cellular component; cell proliferation, e.g., synthesis of DNA; cell migration; cell differentiation; and cell survival

The response mediated by the receptor protein depends on the type of cell it is expressed on. Some information regarding the types of cells that express other members of the subfamily of GPCRs of the present invention is already known in the art (see references cited in Background and  
25 information regarding closest homologous protein provided in Figure 2 and expression information provided in Figure 1: the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening). For example, in some cells, binding of a ligand to the receptor protein may stimulate an activity such as release of compounds, gating of a channel, cellular adhesion, migration, differentiation, etc., through phosphatidylinositol or cyclic AMP  
30 metabolism and turnover while in other cells, the binding of the ligand will produce a different result. Regardless of the cellular activity/response modulated by the particular GPCR of the present invention, a skilled artisan will clearly know that the receptor protein is a GPCR and interacts with G proteins to produce one or more secondary signals, in a variety of intracellular signal transduction pathways, e.g., through phosphatidylinositol or cyclic AMP metabolism and turnover, in a cell thus participating in a

biological process in the cells or tissues that express the GPCR (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening).

As used herein, "phosphatidylinositol turnover and metabolism" refers to the molecules involved in the turnover and metabolism of phosphatidylinositol 4,5-bisphosphate (PIP<sub>2</sub>) as well as to the activities of these molecules. PIP<sub>2</sub> is a phospholipid found in the cytosolic leaflet of the plasma membrane. Binding of ligand to the receptor activates, in some cells, the plasma-membrane enzyme phospholipase C that in turn can hydrolyze PIP<sub>2</sub> to produce 1,2-diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP<sub>3</sub>). Once formed IP<sub>3</sub> can diffuse to the endoplasmic reticulum surface where it can bind an IP<sub>3</sub> receptor, e.g., a calcium channel protein containing an IP<sub>3</sub> binding site. IP<sub>3</sub> binding can induce opening of the channel, allowing calcium ions to be released into the cytoplasm. IP<sub>3</sub> can also be phosphorylated by a specific kinase to form inositol 1,3,4,5-tetraphosphate (IP<sub>4</sub>), a molecule that can cause calcium entry into the cytoplasm from the extracellular medium. IP<sub>3</sub> and IP<sub>4</sub> can subsequently be hydrolyzed very rapidly to the inactive products inositol 1,4-bisphosphate (IP<sub>2</sub>) and inositol 1,3,4-triphosphate, respectively. These inactive products can be recycled by the cell to synthesize PIP<sub>2</sub>. The other second messenger produced by the hydrolysis of PIP<sub>2</sub>, namely 1,2-diacylglycerol (DAG), remains in the cell membrane where it can serve to activate the enzyme protein kinase C. Protein kinase C is usually found soluble in the cytoplasm of the cell, but upon an increase in the intracellular calcium concentration, this enzyme can move to the plasma membrane where it can be activated by DAG. The activation of protein kinase C in different cells results in various cellular responses such as the phosphorylation of glycogen synthase, or the phosphorylation of various transcription factors, e.g., NF-kB. The language "phosphatidylinositol activity", as used herein, refers to an activity of PIP<sub>2</sub> or one of its metabolites.

Another signaling pathway in which the receptor may participate is the cAMP turnover pathway. As used herein, "cyclic AMP turnover and metabolism" refers to the molecules involved in the turnover and metabolism of cyclic AMP (cAMP) as well as to the activities of these molecules. Cyclic AMP is a second messenger produced in response to ligand-induced stimulation of certain G protein coupled receptors. In the cAMP signaling pathway, binding of a ligand to a GPCR can lead to the activation of the enzyme adenylyl cyclase, which catalyzes the synthesis of cAMP. The newly synthesized cAMP can in turn activate a cAMP-dependent protein kinase. This activated kinase can phosphorylate a voltage-gated potassium channel protein, or an associated protein, and lead to the inability of the potassium channel to open during an action potential. The inability of the potassium channel to open results in a decrease in the outward flow of potassium,

which normally repolarizes the membrane of a neuron, leading to prolonged membrane depolarization.

By targeting an agent to modulate a GPCR, the signaling activity and biological process mediated by the receptor can be agonized or antagonized in specific cells and tissues (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening). Such agonism and antagonism serves as a basis for modulating a biological activity in a therapeutic context (mammalian therapy) or toxic context (anti-cell therapy, e.g. anti-cancer agent).

## DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of a cDNA molecule or transcript sequence that encodes the GPCR of the present invention. In addition structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of inventions based on this molecular sequence (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening).

FIGURE 2 provides the predicted amino acid sequence of the GPCR of the present invention. In addition structure and functional information, such as protein family and function, modification sites, is provided that allows one to readily determine specific uses of inventions based on this molecular sequence as well as significant fragments of the proteins of the present invention.

FIGURE 3 provides genomic sequences that span the gene encoding the GPCR protein of the present invention. In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided that allows one to readily determine specific uses of inventions based on this molecular sequence as well as important fragments for use in probe and primer design and heterologous gene expression control. Figure 3 also provides SNP information that has been found in the gene encoding the GPCR proteins of the present invention. The following variations were seen: C8692T, A8716G and A8970G (S to N coding change).

## DETAILED DESCRIPTION OF THE INVENTION

### General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share structural

and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a GPCR protein or part of a GPCR protein, that are related to the MAS proto-oncogene subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human GPCR peptides and proteins, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these GPCR peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the GPCR of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known GPCR proteins of the MAS proto-oncogene subfamily and the expression pattern observed (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening). The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the know MAS proto-oncogene family or subfamily of GPCR proteins.

#### Specific Embodiments

##### Peptide Molecules

The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the GPCR family of proteins (protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figures 1 and genomic sequences are provided in Figure 3). The peptide sequences provided in Figure 2, as well as the obvious variants described herein, such as allelic variants, will be referred herein as the GPCR peptides of the present invention, GPCR peptides, or peptides/proteins of the present invention.

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of or are comprised of the amino acid sequences of the GPCR peptides disclosed in the Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA and



Figure 3, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the GPCR peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated GPCR peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening). For example, a nucleic acid molecule encoding the GPCR peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequences that such a protein consists of is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that are comprised of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein is comprised of an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the GPCR peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The GPCR peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a GPCR peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the GPCR peptide. "Operatively linked" indicates that the GPCR peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the GPCR peptide.

In some uses, the fusion protein does not affect the activity of the GPCR peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant GPCR peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification

of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A GPCR peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the GPCR peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art know techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the GPCR peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular*

*Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)) (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the GPCR peptides of the present invention as well as being encoded by the same genetic locus as the GPCR peptide provided herein. The GPCR of the present invention is encoded by a gene on chromosome 11 near markers SHGC-34534 (LOD=11.64) and SHGC-5996 (LOD=11.64).

Allelic variants of a GPCR peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the GPCR peptide as well as being encoded by the same genetic locus as the GPCR peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in Figure 3, such as the genomic sequence mapped to the reference human (The GPCR of the present invention is encoded by a gene on chromosome 11 near markers SHGC-34534 (LOD=11.64) and SHGC-5996 (LOD=11.64)). As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Figure 3 provides SNP information that has been found in the gene encoding the GPCR proteins of the present invention. The following variations were seen: C8692T, A8716G and A8970G (S to N coding change).

Paralogs of a GPCR peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the GPCR peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a GPCR peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the GPCR peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the GPCR peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the GPCR peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a GPCR peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala,

Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

Variant GPCR peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind ligand, ability to bind G-protein, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis that identifies critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)), particularly using the results provided in Figure 2. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as ligand/effector molecule binding or in assays such as an *in vitro* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al.* *Science* 255:306-312 (1992)).

The present invention further provides fragments of the GPCR peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly fragments identified in Figure 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16 or more contiguous amino acid residues from a GPCR peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the GPCR peptide or could be chosen for the ability to perform a function, e.g. ability to bind ligand or effector molecule or act as an immunogen. Particularly important fragments are biologically active fragments, peptides which are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the GPCR peptide, e.g., active site, a G-protein binding site, a transmembrane domain or a ligand-binding

domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well-known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in Figure 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in GPCR peptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Accordingly, the GPCR peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature GPCR peptide is fused with another compound, such as a compound to increase the half-life of the GPCR peptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature GPCR peptide, such as a leader or secretory sequence or a sequence for purification of the mature GPCR peptide or a pro-protein sequence.

Such modifications are well-known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of*

*Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol.* 182: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

#### Protein/Peptide Uses

5       The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures and Back Ground Section; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or receptor) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either  
10       constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the binding partner so as to develop a system to identify inhibitors of the binding interaction. Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as commercial products.

15       Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

20       The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, GPCRs isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug to modulate the cells or tissues that express the receptor (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary  
25       gland as confirmed by cDNA panel screening). Approximately 70% of all pharmaceutical agents modulate the activity of a GPCR. A combination of the invertebrate and mammalian ortholog can be used in selective screening methods to find agents specific for invertebrates. The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention. Such uses can readily be determined using the  
30       information provided herein, that known in the art and routine experimentation.

      The receptor polypeptides (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to GPCRs. Such assays involve any of the known GPCR functions or activities or properties useful for diagnosis and treatment of GPCR-related conditions that are specific for the subfamily of GPCRs that the one of the present invention



belongs to, particularly in cells and tissues that express this receptor (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening).

5 The receptor polypeptides are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the receptor protein, as a biopsy or expanded in cell culture (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening). In one embodiment, however, cell-based assays involve recombinant host cells expressing the receptor protein.

10 The polypeptides can be used to identify compounds that modulate receptor activity of the protein in its natural state, or an altered form that causes a specific disease or pathology associated with the receptor. Both the GPCRs of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the receptor. These compounds can be further screened against a functional receptor to determine the effect of the compound on the receptor activity. Further, these compounds can be tested in animal or invertebrate  
15 systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the receptor to a desired degree.

Further, the receptor polypeptides can be used to screen a compound for the ability to stimulate or inhibit interaction between the receptor protein and a molecule that normally interacts with the receptor protein, e.g. a ligand or a component of the signal pathway that the receptor protein normally  
20 interacts (for example, a G-protein or other interactor involved in cAMP or phosphatidylinositol turnover and/or adenylate cyclase, or phospholipase C activation). Such assays typically include the steps of combining the receptor protein with a candidate compound under conditions that allow the receptor protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction  
25 with the receptor protein and the target, such as any of the associated effects of signal transduction such as G-protein phosphorylation, cAMP or phosphatidylinositol turnover, and adenylate cyclase or phospholipase C activation.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature* 354:82-84  
30 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell* 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')<sub>2</sub>, Fab expression library fragments, and epitope-

binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the receptor that competes for ligand binding. Other candidate compounds include mutant receptors or appropriate fragments containing  
5 mutations that affect receptor function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) receptor activity. The assays typically involve an assay of events in the signal  
10 transduction pathway that indicate receptor activity. Thus, a cellular process such as proliferation, the expression of genes that are up- or down-regulated in response to the receptor protein dependent signal cascade, can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase.

Any of the biological or biochemical functions mediated by the receptor can be used as an  
15 endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the receptor can be assayed (the GPCR of the present invention is expressed in  
20 the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening).

Binding and/or activating compounds can also be screened by using chimeric receptor proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by  
25 heterologous domains or subregions. For example, a G-protein-binding region can be used that interacts with a different G-protein than that which is recognized by the native receptor. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. Alternatively, the entire transmembrane portion or subregions (such as transmembrane segments or intracellular or extracellular loops) can be replaced with the entire transmembrane portion or  
30 subregions specific to a host cell that is different from the host cell from which the amino terminal extracellular domain and/or the G-protein-binding region are derived. This allows for assays to be performed in other than the specific host cell from which the receptor is derived. Alternatively, the amino terminal extracellular domain (and/or other ligand-binding regions) could be replaced by a domain (and/or other binding region) binding a different ligand, thus, providing an assay for test

compounds that interact with the heterologous amino terminal extracellular domain (or region) but still cause signal transduction. Finally, activation can be detected by a reporter gene containing an easily detectable coding region operably linked to a transcriptional regulatory sequence that is part of the native signal transduction pathway.

5       The receptor polypeptides are also useful in competition binding assays in methods designed to discover compounds that interact with the receptor. Thus, a compound is exposed to a receptor polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble receptor polypeptide is also added to the mixture. If the test compound interacts with the soluble receptor polypeptide, it decreases the amount of complex formed or activity from the  
10   receptor target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the receptor. Thus, the soluble polypeptide that competes with the target receptor region is designed to contain peptide sequences corresponding to the region of interest.

      To perform cell free drug screening assays, it is sometimes desirable to immobilize either the receptor protein, or fragment, or its target molecule to facilitate separation of complexes from  
15   uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

      Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre  
20   plates, which are then combined with the cell lysates (e.g., <sup>35</sup>S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated  
25   by SDS-PAGE, and the level of receptor-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein  
30   trapped in the wells by antibody conjugation. Preparations of a receptor-binding protein and a candidate compound are incubated in the receptor protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the receptor protein target molecule, or which are reactive with receptor

protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the GPCRs of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell  
5 free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

Modulators of receptor protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the receptor pathway, by treating cells or tissues that express the GPCR (the GPCR of the present invention is expressed in the uterus, testis, heart, liver  
10 and mammary gland as confirmed by cDNA panel screening). These methods of treatment include the steps of administering a modulator of the GPCR's activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

In yet another aspect of the invention, the GPCR proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993)  
15 *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with the GPCR and are involved in GPCR activity. Such GPCR-binding proteins are also likely to be involved in the propagation of signals by the GPCR proteins or GPCR targets as, for example, downstream elements of a GPCR-  
20 mediated signaling pathway. Alternatively, such GPCR-binding proteins are likely to be GPCR inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a GPCR protein is fused to a gene  
25 encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a GPCR-dependent complex, the DNA-binding and activation domains of the transcription factor are  
30 brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the GPCR protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a GPCR modulating agent, an antisense GPCR nucleic acid molecule, a GPCR-specific  
5 antibody, or a GPCR-binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or insect model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

10 The GPCR proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening). The method involves contacting a  
15 biological sample with a compound capable of interacting with the receptor protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as  
20 well as tissues, cells and fluids present within a subject.

The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a  
25 genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered receptor activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques  
30 useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

*In vitro* techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagents, such as an antibody or protein binding agent.. Alternatively, the peptide can be detected *in vivo* in a

subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the receptor protein in which one or more of the receptor functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and receptor activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening.). Accordingly, methods for treatment include the use of the GPCR protein or fragments.

### Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or F(ab')<sub>2</sub>, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, Antibodies, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods.

Antibodies are preferably prepared from regions or discrete fragments of the GPCR proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or receptor/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various

enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include  
5 umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

#### 10 Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present  
15 invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition (the GPCR of the  
20 present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening). Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function,  
25 particularly in cells and tissues that express the receptor (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening). When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. If a disorder is characterized by a specific mutation in the protein,  
30 antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening). The diagnostic uses



can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

5        Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

10        The antibodies are also useful for tissue typing (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening). Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

15        The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the GPCR peptide to a binding partner such as a ligand. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information  
20 relating to the proteins of the present invention.

25        The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for antibody arrays.

#### Nucleic Acid Molecules

30        The present invention further provides isolated nucleic acid molecules that encode a GPCR peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the GPCR peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figures 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule, for example from about 1-300 additional nucleotides.

The present invention further provides nucleic acid molecules that are comprised of the nucleotide sequences shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ

ID NO:2. A nucleic acid molecule is comprised of a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprises several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, human genomic sequences (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the GPCR peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the GPCR proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify gene modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in Figure 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could be at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a

peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene (The GPCR of the present invention is encoded by a gene on chromosome 11 near markers SHGC-34534 (LOD=11.64) and SHGC-5996 (LOD=11.64)).

Figure 3 provides SNP information that has been found in the gene encoding the GPCR proteins of the present invention. The following variations were seen: C8692T, A8716G and A8970G (S to N coding change).

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

#### Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2 (Figure 3 provides SNP information that has been found in the gene encoding the GPCR proteins of the present invention. The following variations were seen: C8692T, A8716G and A8970G (S to N coding change)).

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also  
5 include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

10 The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods (The GPCR of the present invention is encoded by a gene on chromosome 11 near markers SHGC-34534 (LOD=11.64) and SHGC-5996 (LOD=11.64)). This is particularly useful in determining whether a particular protein is an allelic variant of one the proteins provided herein

15 The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention as described in detail below.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

20 The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

25 The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening). The nucleic acid whose level is determined  
30 can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in GPCR protein expression relative to normal results.

*In vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a  
5 GPCR protein, such as by measuring a level of a receptor-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a receptor gene has been mutated (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening).

Nucleic acid expression assays are useful for drug screening to identify compounds that  
10 modulate GPCR nucleic acid expression, particularly in cells and tissues that express the receptor (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening).

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the GPCR gene. The method typically includes  
15 assaying the ability of the compound to modulate the expression of the GPCR nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired GPCR nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the GPCR nucleic acid (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel  
20 screening.) or recombinant cells genetically engineered to express specific nucleic acid sequences.

The assay for GPCR nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the GPCR protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked  
25 to a reporter gene such as luciferase.

Thus, modulators of GPCR gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of GPCR mRNA in the presence of the candidate compound is compared to the level of expression of GPCR mRNA in the absence of the candidate compound. The candidate compound can  
30 then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When

nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate GPCR nucleic acid  
5 expression, particularly to modulate activities within a cell or tissue that expresses the proteins (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening). Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

Alternatively, a modulator for GPCR nucleic acid expression can be a small molecule or drug  
10 identified using the screening assays described herein as long as the drug or small molecule inhibits the GPCR nucleic acid expression in the cells and tissues that express the protein (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by cDNA panel screening).

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating  
15 compounds on the expression or activity of the GPCR gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased  
20 administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in GPCR nucleic acid, and particularly in qualitative changes that lead to pathology. The nucleic acid  
25 molecules can be used to detect mutations in GPCR genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally-occurring genetic mutations in the GPCR gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition,  
30 modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the GPCR gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a GPCR protein.



Individuals carrying mutations in the GPCR gene can be detected at the nucleic acid level by a variety of techniques (The GPCR of the present invention is encoded by a gene on chromosome 11 near markers SHGC-34534 (LOD=11.64) and SHGC-5996 (LOD=11.64)). Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis (Figure 3 provides SNP

5 information that has been found in the gene encoding the GPCR proteins of the present invention. The following variations were seen: C8692T, A8716G and A8970G (S to N coding change)). RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR)  
10 (see, e.g., Landegran *et al.*, *Science* 241:1077-1080 (1988); and Nakazawa *et al.*, *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res.* 23:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically  
15 hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA  
20 sequences.

Alternatively, mutations in a GPCR gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched  
25 sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant GPCR gene and a wild-type gene can be determined by direct DNA sequencing. A  
30 variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naevé, C.W., (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science* 230:1242 (1985)); Cotton *et al.*, *PNAS* 85:4397 (1988); Saleeba *et al.*, *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature* 313:495 (1985)). Examples of other techniques for detecting point mutations include, selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the GPCR gene in an individual in order to select an appropriate compound or dosage regimen for treatment. Figure 3 provides SNP information that has been found in the gene encoding the GPCR proteins of the present invention. The following variations were seen: C8692T, A8716G and A8970G (S to N coding change).

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control GPCR gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of GPCR protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into GPCR protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of GPCR nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired GPCR nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include

coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the GPCR protein, such as ligand binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in GPCR gene expression. Thus, recombinant cells, which include the patient's cells  
5 that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired GPCR protein to treat the individual.

The invention also encompasses kits for detecting the presence of a GPCR nucleic acid in a biological sample, particularly cells and tissues that normally express the protein (the GPCR of the present invention is expressed in the uterus, testis, heart, liver and mammary gland as confirmed by  
10 cDNA panel screening). For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting GPCR nucleic acid in a biological sample; means for determining the amount of GPCR nucleic acid in the sample; and means for comparing the amount of GPCR nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect GPCR protein  
15 mRNA or DNA.

#### Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in  
20 Figures 1 and 3 (SEQ ID NOS:1 and 3).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee et al., PCT application  
25 W095/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown et. al., US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-  
30 stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain

oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

5 In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary  
10 structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a  
15 light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by  
20 reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and  
25 may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the  
30 presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The

scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the  
5 absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of the GPCR proteins/peptides of the present invention and allelic variation within this gene/protein. In  
10 detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes or alleles, at least one of which is a gene and or alleles of the GPCR gene of the present invention (Figure 3 provides SNP information that has been found in the gene encoding the GPCR proteins of the present invention. The  
15 following variations were seen: C8692T, A8716G and A8970G (S to N coding change)).

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can  
20 readily be adapted to employ the novel fragments of the Human genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory*  
25 *Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed.  
30 Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the GPCR disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified GPCR genes of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

#### Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host

cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage  $\lambda$ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, eg. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by

temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

5 The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

10 The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein.  
15 Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes  
20 include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene* 69:301-315 (1988))  
25 and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of  
30 interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res.* 20:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*,



*EMBO J.* 6:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell* 30:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology* 170:31-39 (1989)).

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature* 329:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*. 2nd, ed, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one  
5 vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective,  
10 replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate  
15 reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs  
20 described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as GPCRs, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

25 Where the peptide is not secreted into the medium, which is typically the case with GPCRs, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography,  
30 hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the

cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

#### Uses of vectors and host cells

5       The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a GPCR protein or peptide that can be further purified to produce desired amounts of GPCR protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

10       Host cells are also useful for conducting cell-based assays involving the GPCR protein or GPCR protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native GPCR protein is useful for assaying compounds that stimulate or inhibit GPCR protein function.

15       Host cells are also useful for identifying GPCR protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant GPCR protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native GPCR protein.

20       Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a GPCR protein and identifying and evaluating modulators of GPCR protein activity. Other examples of transgenic animals include non-  
25       human primates, sheep, dogs, cows, goats, chickens, and amphibians.

      A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the GPCR protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

30       Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the GPCR protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes.

5     A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS* 89:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science* 251:1351-1355 (1991). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

15     

20     

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

25     

30     

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, GPCR protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays.

Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* GPCR protein function, including ligand interaction, the effect of specific mutant GPCR proteins on GPCR protein function and ligand interaction, and the effect of chimeric GPCR proteins. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more

5 GPCR protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred  
10 embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

## SEQUENCE LISTING

&lt;110&gt; Wei et al.

<120> ISOLATED HUMAN G-PROTEIN COUPLED  
RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR  
PROTEINS, AND USES THEREOF

&lt;130&gt; CL000751PCT

&lt;140&gt; N/A

&lt;141&gt; 2001-03-27

&lt;150&gt; US 60/192,853

&lt;151&gt; 2000-03-29

&lt;150&gt; US 09/635,593

&lt;151&gt; 2000-08-09

&lt;160&gt; 7

&lt;170&gt; FastSEQ for Windows Version 4.0

&lt;210&gt; 1

&lt;211&gt; 993

&lt;212&gt; DNA

&lt;213&gt; Human

&lt;400&gt; 1

```

atggatccaa ccaccccggc ctggggaaca gaaagtacaa cagtgaatgg aaatgaccaa 60
gcccttcttc tgccttgtgg caaggagacc ctgatccggg tcttcctgat ccttttcatt 120
gccctgggtcg ggctggtagg aaacgggttt gtgctctggc tcctgggctt ccgcatgcgc 180
aggaacgcct tctctgtcta cgtcctcage ctggccgggg ccgacttcct ctctctctgc 240
ttccagatta taaattgcct ggtgtacctc agtaacttct tctgttccat ctccatcaat 300
ttccctagct tcttcaccac tgtgatgacc tgtgcctacc ttgcaggcct gagcatgctg 360
agcacccgtca gcacccgagcg ctgcctgtcc gtctctgtgg ccatctggta tcgctgcgcg 420
cgccccagac acctgtcagc ggtcgtgtgt gtctctgtct gggccctgtc cctactgctg 480
agcatcttgg aagggaagt ctgtggcttc ttatttagtg atggtgactc tggttggtgt 540
cagacatttg atttcatcac tgcagcgtgg ctgatttttt tattcatggt tctctgtggg 600
tccagtctgg cctgtctggt caggatcctc tgtggctcca ggggtctgcc actgaccagg 660
ctgtacctga ccatcctgct cacagtgtcg gtgttctctc tctgcggcct gccctttggc 720
attcagtggt tcctaataatt atggatctgg aaggattctg atgtcttatt ttgtcatatt 780
catccagttt cagttgtcct gtcactcttt aacagcagtg ccaaccccat catttacttc 840
ttcgtgggct cttttaggaa gcagtggcgg ctgcagcagc cgatcctcaa gctggctctc 900
cagagggctc tgcaggacat tgcagagtg gatcacagtg aaggatgctt ccgtcagggc 960
accccgagaga tgcagagaag cagtctggtg tag 993

```

&lt;210&gt; 2

&lt;211&gt; 330

&lt;212&gt; PRT

&lt;213&gt; Human

&lt;400&gt; 2

```

Met Asp Pro Thr Thr Pro Ala Trp Gly Thr Glu Ser Thr Thr Val Asn
 1           5           10          15
Gly Asn Asp Gln Ala Leu Leu Leu Cys Gly Lys Glu Thr Leu Ile
 20          25          30
Pro Val Phe Leu Ile Leu Phe Ile Ala Leu Val Gly Leu Val Gly Asn
 35          40          45
Gly Phe Val Leu Trp Leu Leu Gly Phe Arg Met Arg Arg Asn Ala Phe
 50          55          60
Ser Val Tyr Val Leu Ser Leu Ala Gly Ala Asp Phe Leu Phe Leu Cys
 65          70          75          80

```

aaaaatcctc	aataaagtac	tgtcaaaccg	aatccagcag	cacatccaaa	agtttatcca	60
ccatgatcaa	gtcggcttca	tccctgggat	gcaagtctgg	ctcaaaatac	gcaaatcaat	120
aatgtaatc	catcacataa	gcagaaccaa	tgacaaaaac	cacatgatta	tctcaataga	180
tgcagaaaaa	gctgtcgtata	aaattcaaca	ccccctcatg	ctaaaaactc	tcaataaaact	240
gggtactgat	ggaacatatc	tcaaaataat	aagagctatt	ttttataaac	ccacagcaaa	300
tatcacactg	aatgggcaaa	agctggaaga	attcccttgg	aaaattgccca	caagacaagg	360
atgccctctc	tcaccactcc	tattcaacat	agtattggaa	gttctgcaca	gagaaatcag	420
acaagaaaaa	gaaataaagg	gtattcaatt	aggaaaagag	gacgtcaaat	tgtccctgtt	480
tgcagatgac	atgattgtat	atttagaaaa	ctccattgtc	tcagtcccaa	atctccttaa	540
gctgataaag	aactctcaga	aagctctcag	atacaaaaatc	aatgtgcaaa	aatcacaagc	600
attcccatc	accaataata	gacaaacaga	gagccaaatc	atgagtgaac	tctcatttac	660
aattgctaca	aagagaataa	aatacctagc	aatacaattt	acaaaggatg	tgaaggacct	720
cttcaaggag	aactacaaac	cactgctcaa	ggaaataaga	gaggatacaa	acaaatggaa	780
aaaaattcca	tgctcatgga	taggaagaat	caatgtcgtg	aaaatggcca	tactgcccaa	840
agtaatttca	agttttaaag	ctatccacgt	caggctacca	ttgactttct	tcacagaagt	900
agaaaaaact	acttttaaatt	tcatatggaa	acaaaagaga	tcctgtatag	ccaggacaat	960
cctaagcaaa	aagaataaag	ctggaggcaa	cacactacct	gacttcaaac	tatactacaa	1020
ggctacagta	accaaaacaa	catgggtgctg	gtactaaaac	atataatatcg	accagtggaa	1080
cagaacagag	gcctcagaaa	tgatgccaca	catgtacaac	catctgatct	ttgacaaacc	1140
tgacaaaaac	aagcaatggg	gaaaggattc	cttattttaa	aaatgtgtgt	gggaaaaactg	1200
gttagccata	tcgagaaaaa	tgaactgga	cccccttctt	acaccttata	caaaaatttaa	1260
ctcaaaatgg	attaaagact	taaacataaag	acctaaaacc	ataaaacctc	aqaaqaaaac	1320

ctagacagta	ccattcagga	cataggcattg	ggcaaagacc	tcatgaataa	aacacccaaa	1380
gcaatggcaa	caaaagccaa	aattgacaaa	tgggatctaa	ttaaactaaa	gagcttctgc	1440
acagcaaaag	aaactatcat	cagagtgaac	aggcaaccta	cagaatgaga	gaaaaatttt	1500
gcaatctatc	catctgacaa	agggctaata	tccagaattt	acaagaaact	taaacaaaat	1560
tataagaaaa	aataaacaac	tccatcaaaa	agtgggtgaa	gaatatgaac	agagacttct	1620
caaaagaaga	tatttatatg	gccaaaccaac	atgaaaaaat	gctcatcatc	actggtcatt	1680
agaggaatgc	aaattaaaaac	cacaatgaga	taccatctta	caccagttag	aatggaaatc	1740
attaaaaagt	caggaaacaa	cagatgctgg	agaggatgtg	gagaaatagg	aacgctttta	1800
cacggttggt	gggagtgcac	attagttcaa	ccattgtgga	agacagtgtg	gcgattccct	1860
aaggatctag	aaccagaaat	accattttgtc	tcagcaatcc	cactactggg	tatatacca	1920
agagattata	aatcattcta	ctataaagac	acatgcacac	atacgtttat	tgcagcacta	1980
ttaacaatag	caaagacctg	gaaccaaccc	aaatgcccat	caatgataga	ctggataaag	2040
aaaatgtggt	acatatcac	catggaattc	tatgcagccg	taagaaaagg	atgagtttat	2100
gttctttgca	gggacgtgga	tgaagctgga	aaccatcatt	ctcagcaaac	taacacagga	2160
acagaaaaac	aaacactgca	tggtctcact	ccatgtgtgg	agttgaacaa	tgagaacaca	2220
tgagatacag	gaggggaaca	tcacacactg	gggcctctca	gggggtgggg	agctaggaga	2280
gggatagcac	taggagaaat	acctaagtga	gatgacgggt	tgatggatgc	agcaaacacc	2340
cacggcatgt	gtatacgtag	gtaacaaacc	tgacagttct	gcccagtgtat	cgcataactt	2400
aaagtataat	tttaaaaaaa	taaaatagaa	cttctagaaa	ttagtaatat	aaaaattgga	2460
tttaaaacttc	agtggataga	ttaaacagac	gattagacac	aatggaaaga	atctagttaa	2520
ctggaagata	catctgaaga	aattaccag	aaagcagcct	cagagacaga	gacgagacat	2580
aaggaaaaag	ttaaagaaat	aaggatgata	gaagaagagg	gctaactgtat	gttgaattag	2640
acttcagaa	agagataata	ctaaaagtga	taatgagagc	taataatttt	atagtactta	2700
ctatgtgctc	ctattgtaaa	tcattttatat	acattaaatt	attaatcttc	acagcaaac	2760
tatagattaa	taatgttacc	cccattttac	tgatgaggaa	acaaacataa	cagagttaca	2820
aaatctggct	aaggtcccat	aggagaaatg	ccaagttagc	tggtcttgtg	ttcttcaccg	2880
ttatgttaat	atttaaagag	attatagcat	taaatgtttt	aaaaatgtta	aatgatgata	2940
attttaaaat	ctaggaaact	ccacaaatcc	caaacatgat	ttgaaaaaga	gtcaccagat	3000
acaatttggt	gagaatgtat	catagcagag	aaaaataaaa	gctcttagag	gaaaaaaagg	3060
cagatcacca	caagcaaatg	gcaacctgac	tgatggccac	aacactggca	gtcagaacat	3120
ggatcttcaa	tgtgctggaa	gaaagaaact	atcacacctac	aatgccatgc	ctggaaaaac	3180
catctttaga	agacgagggt	gaaacaatga	cattttcaga	tagataaaag	gaggcattta	3240
ccatgcagag	actotcacag	aaggaaactcc	tgaagaatgt	gttttgagaa	atggaagata	3300
tgagatgtaa	aatgggagga	tgagcaaaag	cattggtgat	tatgtggata	aatctgaata	3360
aacactgact	atttaaaaca	ataataagtg	tttcatttgt	tggagggaaa	cagaatcaaa	3420
atcaaataat	agagcacagt	aggttagaca	gtaacggaag	ttaaagcttt	ctaaggctct	3480
gctgctgctc	atgaggaaag	cagatttctg	gggacaaatg	catggtttac	ccactgcacc	3540
aaagtataat	acaagaaaaat	gctgagtctc	agaactcgcc	tcaatcaaca	catgttcctt	3600
gtctgagggtc	accagctgc	tctggaatgc	aaaggggcat	gatcttgtat	ggtgttcac	3660
attgtgtctt	cattgcctgg	aacggtacct	gacatatcat	aggcaaacaa	taaacagtgt	3720
tgaatgaatg	aataaatgag	tgaataattt	cttctgattg	tcttaaatgc	ttcataattg	3780
ttacttaata	ctcaatttgg	ctgcagcata	ggcacatgaa	aggaagtcac	aggagataag	3840
gctgataata	tgggttggtc	ctagatgatc	taaaaaataat	ctagctaata	aatatcatga	3900
tgtagaggga	gaaacgtgac	agatggtgca	tgtgatgtgc	taagtccctc	atattttata	3960
aggagactca	atatacttag	ttgttaacat	tgataaatca	aaaatagaaa	tggtgtttta	4020
ttacgcaaaag	cttttttaaa	aaaaagaatg	gaaactgtat	gcagtgggtg	cttttgggga	4080
atgagacttc	gagagacttt	gcatttcatt	ttatacacct	tagtgtcatt	tgaataataac	4140
ctgattttga	aaggtcaagt	aagccagggc	cttgaatgcc	aagcagaaga	gggcagacta	4200
ttcttttagct	atgaaagctc	acgtcgcttg	attggtctgc	ttcctggatg	ctcattgtgg	4260
gcaatgaaag	acttcaggac	ggtgattcct	tgacgggagt	ggggccggag	agggtagggg	4320
agaaaccaag	ataagtgatt	ggagttaggg	caaagtctga	gaaacatctc	atgtggtttc	4380
ctgtgtcata	ctcttatttt	ccagcacctt	ctgtacagga	aatcacaggg	ttgttgagca	4440
tcttgataag	cttcactcag	tcactcaggt	tcactccaga	tcctgcaggc	atctccccat	4500
cctcagctgt	tggcagtc	caggaaagca	cttctcaact	caccaactcc	agtagaaaga	4560
agggtgttaa	ggtaagaggt	tggtcaagaa	ccatcttctt	tcaaaggcag	ttttggtacc	4620
ttagtccaca	tcctgatgaa	ttcattagga	aaagaagaca	ggaagctcct	ttgagggaagc	4680
tttggatgct	ggcagcttag	atggtgggag	atccgaatgg	ggtcaagcag	gttcgtccca	4740
cttctcttct	atctaccaat	attctccctt	tctcttcgac	cattcaaate	tgaactatg	4800
gagagaaaga	ttttgctgat	tttaatttag	gggcaggcta	gaatgaagtc	attactccc	4860
taagctaaat	tggtatttaa	aacctcaggt	tgactgatt	ttcttggtat	tcaaaagggt	4920
tttcttctat	atctgtcatt	gtcccagcag	aaaacacaca	tcacacactc	taaaggggta	4980
gtttagagga	gtttagtgaa	gagactgttt	acaacggcgt	gaacagggat	taagggaatga	5040
tgaagccctc	aaaggcaagt	gatagtggag	ggctgttacc	gttgctagggt	ctgaagtgtat	5100



aaggagaaag	agaagttctg	agaactcaga	atgtgctgta	gcggtaactg	taactgtaac	5160
tatagctgtg	gctacaggaa	agggccacca	gacagggcta	tgtccttagg	tagaaaaaca	5220
ctgccactgc	caactcacag	cccttcaggg	cgaggaggga	gagccaggaa	atTTTTTaa	5280
aatcatcctc	caatctcctg	tcaatgtgtc	cccttggtg	aaaaaaaaag	tcacctoca	5340
atctcctgtc	aatgtgtacc	ctttggagcc	tgagtgaag	acagccatt	gacgaggcac	5400
agacatgtct	cctcccagga	tgcaaaggca	ggtagagaag	gatggacatg	agatccta	5460
agcaaataag	aaaagtccat	ttcttcccat	agcctcttct	cagtttgtct	ttcctgagac	5520
tttctctatt	aatgtgattg	aatcaatttc	tcattctatc	acctcccttt	TTTTTaattg	5580
tttgaccttc	ccatctagt	tcacttcttt	ggactagtct	ctcactatca	tcataaatgc	5640
cttgagaatg	gaatgtggtt	gggaaaaaaa	gggattggga	gtacataggt	actcccagct	5700
ataagtacac	aggatgtgt	aatgtaagtt	tatgattttg	gcttctctaa	aaagaaacct	5760
attatagtga	taataaaaa	gaatgtttat	tagacttggg	ctgataactc	taagacatca	5820
aacatcaagg	ctccttccaa	tttctatttc	cccatcccta	aagcgttacc	cttaccctca	5880
tggctctaca	tggctcccaa	acacatcagc	attccagcca	gagagaagag	aaaaaaaaag	5940
agagtacagg	aggagctggt	tatatcactt	tctctcacct	ccccttgacc	atcacatatg	6000
gttcatggtt	gccattagca	taaaagggaag	ctgagaaatg	ggcctttttc	tgatctaaga	6060
ctcagaatta	ttatgtgtga	gaaaaggaga	atggcttttg	gggcaacaag	cagtctctat	6120
agcccgatlc	tctttttagca	tctgtgtctca	cctggagatt	tttctcaaca	ctgcctcaca	6180
aatagtgaat	gatgtagaca	tggaacagaa	ataaacctag	gacagggccc	caggagactg	6240
gagctggtat	tggacctgct	tttcaocatg	tgtatcagga	aactcttttg	tacatgttaa	6300
gctgtctctt	ctttgtttaa	tgagcgtatt	acaatagatg	gcctccaagg	gatatgtgtt	6360
ggaggctggg	gatataatag	tgaacaacca	caacaaaaat	gttgagctta	atgagggaag	6420
taaaagaaaa	aacagacata	cacttaaaaa	tacaggtgca	gggccgggca	cggtggctca	6480
cgctctgaat	cctggcactt	tgggaggccg	aggcgggcag	atcacaaagt	cagaagtctg	6540
agaccatcct	ggctaacaca	gtgaaacccc	gtctctactg	aaaaaaaaaa	tacaaaaaat	6600
tagccagtcg	tgggtggggg	cacctgtagt	cccagctact	cgggaggctg	aggcaggaga	6660
atggcgtgag	cccaggaggc	ggagcttgca	gtgagccgag	atcgagccac	tgactccag	6720
cctgggtgac	agagcgagac	tccatctcaa	aaaaaaaaaa	aaaatacagg	tgcaaaactgt	6780
gttcagtgcc	acgaataaaa	aggccaacct	gcctgagggg	aggacaacag	aagtggtttt	6840
cttccatatt	ctcacaaagg	caaccagggc	tacaggaact	aagacagcat	gtgatggatg	6900
cattcagtc	ttgatgccaa	ttctgcaccc	ccatggagga	cacattcact	ttccccttgg	6960
aggcctccaa	cccatgagag	ttttctgcag	tgtggagtga	ttgacgctg	ttttctgaa	7020
gcaagaaaga	gctcttgcat	tcagttgctt	attctcagtt	ggcttctatt	catactctct	7080
ctccttattt	gattaatctc	tcttccttac	tgtttcttaa	agttgcttaa	tctaggctta	7140
aaaaactgca	atttcgacat	tccatttttg	atttgatatg	agcattactt	gttgcttttc	7200
tcctttcttc	atgtgtagtt	ggtattgaac	ccttgcaatg	gggctgcaag	gctgggttaag	7260
aacaagccct	ctgacacaga	cctgtgggtt	caaatgcagg	ctctaccact	tgtttgtgag	7320
acctgaagca	tattactcaa	tacatttggt	ctttaatctc	cttctctgta	aaatgggag	7380
aataatagta	ctgacctcac	caggctctga	agaagattta	aagaggctat	acctgaggag	7440
cactcagcca	gctgctatta	ttcaccccat	tatgagtaag	ggctccttcc	caagtgtaat	7500
ttgcataaag	atctgccttt	ttcccccaaa	gcctgcaca	gtgtagacat	ggttcttttt	7560
tctaattttc	cactcctttc	cccaaataat	cctgttcaca	aatgctttaa	tttgcaaaaga	7620
aaagctcaaaa	tcatctgggg	gaagttaatg	gtccacctta	gatgtattca	taccacaaca	7680
ccctctagtt	acctctagac	aatagacact	ctcaattgct	gcctctctct	ccttttccca	7740
aattgcccc	ttttcctcca	agatgcagag	aaattcagat	aaaccctgaa	tcctctccat	7800
aaactggaag	tttaaaactct	tttggtggt	gagcaaaaag	catatactag	cccttccctg	7860
attgaagata	tgacagtggc	cattattcac	tatctgtgta	ggcagaagca	atcccttcaa	7920
atacaccaga	gcttttccag	aaggaaacat	gaagaatgat	gttttcttca	tttctttaca	7980
tgtgtgtttt	ttagtgaact	atttcatttt	aaaatgttgg	tgttccactt	agcccaaat	8040
ttcaactact	cttgggtccc	catgcatgca	tgcatgcatt	cattcattca	tttatatgtg	8100
gtacaaactt	cttgaccttc	cactatgccc	tcacattgag	ctaggcactg	aggacacaga	8160
aatggataag	agaaaaactc	tttctgtttt	cataaagctg	acctactgag	caatttctag	8220
cctgacaggc	ctggctgtcc	tcagacctgt	cttagcaaga	agtcacatct	ccgttataac	8280
cctaaatcat	gactctccat	ctgactgttg	atggaccttc	ctcagtcatt	aagggtggcc	8340
atgggtcatg	gcctgggtca	tttcttctgg	ccctgagag	gaaatctgta	tgccaggata	8400
aaaagaatgc	tgaactcaga	gtcctgtctc	ttgctgcccc	acataattctg	cctagacttt	8460
tctcttctctg	cctccaggag	acatgacaca	gtgacagtga	gtgggggtgt	ttgggcctta	8520
gaatattccc	ataccagcag	aggatctacc	ttgtaataga	aagagcccag	gatttagagt	8580
cagcaagaaa	tgagtttgaa	tccagtgct	aggactccct	ggcccttaat	aaatgactta	8640
atctcttcaa	gcctctgatt	tcctctcctg	taaaacaggg	gtggtaatta	ccacataaca	8700
ggctgggtcat	gagaatcagt	gaacatgcag	caggtgctca	agtcttgttt	ttgtttccag	8760
gggcaccagt	ggaggttttc	tgagcatgga	tccaaccacc	ccggcctggg	gaacagaaag	8820
tacaacagtg	aatggaaatg	accaagccct	tcttctgctt	tgtggcaagg	agaccctgat	8880

```

ccccgtcttc ctgacccctt tcattgccct ggtcgggctg gtaggaaacg ggtttgtgct 8940
ctggctcctg ggcttcocga tgcgcaggag cgccttctct gtctacgtcc tcagccctggc 9000
cggggcccgc ttcctcttcc tctgcttcca gattataaat tgccctgggt acctcagtaa 9060
cttctctctgt tccatctcca tcaatttccc tagcttcttc accactgtga tgacctgtgc 9120
ctaccttgca gccctgagca tgctgagcac cgtcagcacc gagcgctgcc tgtccgtcct 9180
gtggcccatc tggatcgct gccgcgcgcc cagacacctg tcagcggtcg tgtgtgtcct 9240
gctctgggccc ctgtccctac tgctgagcat cttggaaggg aagttctgtg gcttcttatt 9300
tagtgatggt gactctggtt ggtgtcagac atttgatttc atcactgcag cgtggctgat 9360
ttttttattc atggttctct gtgggtccag tctggccctg ctggtcagga tcctctgtgg 9420
ctccaggggt ctgccactga ccaggctgta cctgaccatc ctgctcacag tgctgggtgt 9480
cctctctctgc ggctgcctt ttggcattca gtggttctta atattatgga tctggaagga 9540
ttctgatgtc ttattttgtc atattcatcc agtttcagtt gtccgtcatc ctcttaacag 9600
cagtgccaac cccatcattt acttcttctg gggctctttt aggaagcagt ggcggctgca 9660
gcagccgatc ctcaagctgg ctctccagag ggctctgcag gacattgctg aggtggatca 9720
cagtgaagga tgcttccgtc agggcaccct ggagatgtcg agaagcagtc tgggtagag 9780
atggacagcc tctacttcca tcagatatat gtggctttga gaggcaactt tgcccctgtc 9840
tgtctgattt gctgaacttt ctacgtctct attttaaaac agttaagaga gtccttgtga 9900
ggattaagtg agacagtgcc tatgaaacaa acactaagtg cagtgtctct ggaactgcct 9960
tactcacagg ctccaccac agccctatga gagctttgcc aactctgcgg tccatgactg 10020
ttcccacttt taatgaatcc tacctttcgc agaaggtgta aagcagggca gaaaaggctc 10080
acatttcttt ggtcactgca cttgataggg actcaaacaa tgttatattt ttaattaatt 10140
tctttttctc ttccgtacaa tttctgtctc aacaaaatta gaagaattaa atttaaaact 10200
agctccaaaa gagcagtcgt ctttcatttt ggcagacctt agaatatccc cctagcttaa 10260
taaatctttg ttgaatggct taatgaatga ataaactggg taacgtttta gttaaacctc 10320
tgaaaagtct ccatttacca gatttgagtc actaaattta ttgctttcac tacttttgaa 10380
ttttgcaaac atgaaattaa gttttataat tagataaatc aatgtcaaca catatttaa 10440
gtttgaggta cactgtcttc ctgtggtttc ctttcacatg ccattccctta aaatcccagc 10500
tacacgcctt cccattcctt cccctttgcc tttgttctaa tcttccctct ctgggggctc 10560
tctaattcgt cctggaagtt tccaatggtc ttatagactc catgttcttg gaggacaggc 10620
tgtatgtcag atttaccttt tattccgaag aactcggagc atttattttg ttaattaaat 10680
tgcacatatt tttaaaagtt acgtgttcca cagaataaaa tactaattgt aaatgctgca 10740
tcttttaata attttttatt atttttaatt aaggtataat taagagaatg atatgcacag 10800
atcataaatg tacagcttaa tgagttttgt taaatgaatg tgactcaaac ttcattcaag 10860
acaaaatatt ttcacatctc ctaaaagttc ccctgtgctt cctcattctt cctgccccca 10920
gaggcaacca ttgtttctat ttcaatcaac aaatatttagc tttgcctgac ctttaactgt 10980
atctttttag aatcataata tgtgtatttt ttgtttctag ctcttgcaact caacataata 11040
ttttga 11046

```

&lt;210&gt; 4

&lt;211&gt; 271

&lt;212&gt; PRT

&lt;213&gt; Rattus norvegicus

&lt;400&gt; 4

```

Tyr Ile Phe Leu Leu Cys Leu Cys Gly Leu Val Gly Asn Gly Leu
1           5           10           15
Val Leu Trp Phe Phe Gly Phe Ser Ile Lys Arg Thr Pro Phe Ser Ile
20           25           30
Tyr Phe Leu His Leu Ala Ser Ala Asp Gly Ile Tyr Leu Phe Ser Lys
35           40           45
Ala Val Ile Ala Leu Leu Asn Met Gly Thr Phe Leu Gly Ser Phe Pro
50           55           60
Asp Tyr Val Arg Arg Val Ser Arg Ile Val Gly Leu Cys Thr Phe Phe
65           70           75           80
Ala Gly Val Ser Leu Leu Pro Ala Ile Ser Ile Glu Arg Cys Val Ser
85           90           95
Val Ile Phe Pro Met Trp Tyr Trp Arg Arg Arg Pro Lys Arg Leu Ser
100          105          110
Ala Gly Val Cys Ala Leu Leu Trp Leu Leu Ser Phe Leu Val Thr Ser
115          120          125
Ile His Asn Tyr Phe Cys Met Phe Leu Gly His Glu Ala Ser Gly Thr

```

130	135	140
Ala Cys Leu Asn Met Asp Ile Ser Leu Gly Ile Leu Leu Phe Phe Leu		
145	150	155
Phe Cys Pro Leu Met Val Leu Pro Cys Leu Ala Leu Ile Leu His Val		160
	165	170
Glu Cys Arg Ala Arg Arg Arg Gln Arg Ser Ala Lys Leu Asn His Val		175
	180	185
Val Leu Ala Ile Val Ser Val Phe Leu Val Ser Ser Ile Tyr Leu Gly		190
	195	200
Ile Asp Trp Phe Leu Phe Trp Val Phe Gln Ile Pro Ala Pro Phe Pro		205
	210	215
Glu Tyr Val Thr Asp Leu Cys Ile Cys Ile Asn Ser Ser Ala Lys Pro		220
225	230	235
Ile Val Tyr Phe Leu Ala Gly Arg Asp Lys Ser Gln Arg Leu Trp Glu		240
	245	250
Pro Leu Arg Val Val Phe Gln Arg Ala Leu Arg Asp Gly Ala Glu		255
	260	265
		270

<210> 5  
 <211> 60  
 <212> DNA  
 <213> Human

<400> 5  
 aaacaggggy ggtaattacc acataacagg ctggatcatga aaatcagtga acatgcagca 60

<210> 6  
 <211> 91  
 <212> DNA  
 <213> Human

<400> 6  
 aaacaggggc ggtaattacc acataacagg ctggatcatga raatcagtga acatgcagca 60  
 ggtgctcaag tottggtttt gttccagg g 91

<210> 7  
 <211> 101  
 <212> DNA  
 <213> Human

<400> 7  
 ggtaggaac gggtttgtgc tctggctcct gggcttcgc atgcgcagga rcgccttctc 60  
 tgtctacgtc ctcagcctgg ccggggccga cttcctcttc c 101

### Claims

That which is claimed is:

1. An isolated peptide consisting of an amino acid sequence selected from the group consisting of:
  - (a) an amino acid sequence shown in SEQ ID NO:2;
  - (b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
  - (c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic); and
  - (d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.
2. An isolated peptide comprising an amino acid sequence selected from the group consisting of:
  - (a) an amino acid sequence shown in SEQ ID NO:2;
  - (b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
  - (c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic); and
  - (d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.
3. An isolated antibody that selectively binds to a peptide of claim 2.

4. An isolated nucleic acid molecule consisting of a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;
- (b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
- (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
- (d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and
- (e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

5. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;
- (b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
- (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
- (d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and
- (e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

6. A gene chip comprising a nucleic acid molecule of claim 5.

7. A transgenic non-human animal comprising a nucleic acid molecule of claim 5.
8. A nucleic acid vector comprising a nucleic acid molecule of claim 5.
9. A host cell containing the vector of claim 8.
10. A method for producing any of the peptides of claim 1 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
11. A method for producing any of the peptides of claim 2 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
12. A method for detecting the presence of any of the peptides of claim 2 in a sample, said method comprising contacting said sample with a detection agent that specifically allows detection of the presence of the peptide in the sample and then detecting the presence of the peptide.
13. A method for detecting the presence of a nucleic acid molecule of claim 5 in a sample, said method comprising contacting the sample with an oligonucleotide that hybridizes to said nucleic acid molecule under stringent conditions and determining whether the oligonucleotide binds to said nucleic acid molecule in the sample.
14. A method for identifying a modulator of a peptide of claim 2, said method comprising contacting said peptide with an agent and determining if said agent has modulated the function or activity of said peptide.
15. The method of claim 14, wherein said agent is administered to a host cell comprising an expression vector that expresses said peptide.
16. A method for identifying an agent that binds to any of the peptides of claim 2, said method comprising contacting the peptide with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide.

17. A pharmaceutical composition comprising an agent identified by the method of claim 16 and a pharmaceutically acceptable carrier therefor.
18. A method for treating a disease or condition mediated by a human proteases, said method comprising administering to a patient a pharmaceutically effective amount of an agent identified by the method of claim 16.
19. A method for identifying a modulator of the expression of a peptide of claim 2, said method comprising contacting a cell expressing said peptide with an agent, and determining if said agent has modulated the expression of said peptide.
20. An isolated human protease peptide having an amino acid sequence that shares at least 70% homology with an amino acid sequence shown in SEQ ID NO:2.
21. A peptide according to claim 20 that shares at least 90 percent homology with an amino acid sequence shown in SEQ ID NO:2.
22. An isolated nucleic acid molecule encoding a human protease peptide, said nucleic acid molecule sharing at least 80 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic).
23. A nucleic acid molecule according to claim 22 that shares at least 90 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic).

1/8

## FIGURE 1

```

1  ATGGATCCAA CCACCCCGGC CTGGGGAACA GAAAGTACAA CAGTGAATGG
51  AAATGACCAA GCCCTTCTTC TGCTTTGTGG CAAGGAGACC CTGATCCCGG
101 TCTTCCTGAT CCTTTTCATT GCCCTGGTCG GGCTGGTAGG AAACGGGTTT
151 GTGCTCTGGC TCCTGGGCTT CCGCATGCGC AGGAACGCCT TCTCTGTCTA
201 CGTCCTCAGC CTGGCCGGGG CCGACTTCCT CTTCTCTGCG TTCCAGATTA
251 TAAATTGCCT GGTGTACCTC AGTAACTTCT TCTGTTCCAT CTCCATCAAT
301 TTCCCTAGCT TCTTCACCAC TGTGATGACC TGTGCCTACC TTGCAGGCCT
351 GAGCATGCTG AGCACCCTCA GCACCGAGCG CTGCCTGTCC GTCCTGTGGC
401 CCATCTGGTA TCGCTGCCGC CGCCCAGAC ACCTGTGAGC GGTGCTGTGT
451 GTCCTGCTCT GGGCCCTGTC CCTACTGCTG AGCATCTTGG AAGGGAAGTT
501 CTGTGGCTTC TTATTTAGTG ATGGTGACTC TGGTTGGTGT CAGACATTTG
551 ATTTTCATCAC TGCAGCGTGG CTGATTTTTT TATTCATGGT TCTCTGTGGG
601 TCCAGTCTGG CCCTGCTGGT CAGGATCCTC TGTGGCTCCA GGGGTCTGCC
651 ACTGACCAGG CTGTACCTGA CCATCCTGCT CACAGTGCTG GTGTTCCTCC
701 TCTGCGGCCT GCCCTTTGGC ATTCAGTGGT TCCTAATATT ATGGATCTGG
751 AAGGATTCTG ATGTCTTATT TTGTCATATT CATCCAGTTT CAGTTGTCCT
801 GTCATCTCTT AACAGCAGTG CCAACCCCAT CATTTACTTC TTCGTGGGCT
851 CTTTTAGGAA GCAGTGGCGG CTGCAGCAGC CGATCCTCAA GCTGGCTCTC
901 CAGAGGGCTC TGCAGGACAT TGCTGAGGTG GATCACAGTG AAGGATGCTT
951 CCGTCAGGGC ACCCCGGAGA TGTCGAGAAG CAGTCTGGTG TAG .(SEQ ID NO:1)

```

**FEATURES:**

Start: 1  
Stop: 991

**HOMOLOGOUS PROTEIN:**Top BLAST Hits:

gi 134079 sp P23749 RTA_RAT PROBABLE G PROTEIN-COUPLED RECEPTOR...	125	7e-28
gi 6981186 ref NP_036889.1   MAS1 oncogene >gi 135921 sp P12526...	111	1e-23
gi 4505105 ref NP_002368.1   MAS1 oncogene >gi 135920 sp P04201...	109	6e-23
gi 6678804 ref NP_032578.1   MAS1 oncogene >gi 266505 sp P30554...	106	5e-22

**Tissue expression for modulatory use (from cDNA screening panel):**

Human Uterus  
Human Testis  
Human Heart  
Human mammary gland  
Human liver



2/8

## FIGURE 2, sheet 1 of 2

```

1  MDPTTPAWGT ESTTVNGNDQ ALLLLCGKET LIPVFLILFI ALVGLVGNF
51 VLWLLGFRMR RNAFSVYVLS LAGADFLFLC FQIINCLVYL SNFFCSISIN
101 FPSFFTVMVT CAYLAGLSML STVSTERCLS VLWPIWYRCR RPRHLSAVVC
151 VILWALSILL SILEGKFCGF LFSDGDSGWC QTFDFITAAW LIFLEMVLCG
201 SSLALLVRIL CGSRGLPLTR LYLTILLTVL VFLLCGLPFG IQWFLILWIW
251 KDSDVLFCHI HPVSVVLSSL NSSANPIIYF FVGSRKQWR LQQPILKLAL
301 QRALQDIAEV DHSEGCFRQG TPMSRSSLV (SEQ ID NO:2)

```

## FEATURES:

## Functional domains and key regions:

[1] PDOC00001 PS00001 ASN\_GLYCOSYLATION

N-glycosylation site

271-274 NSSA

[2] PDOC00005 PS00005 PKC\_PHOSPHO\_SITE

Protein kinase C phosphorylation site

Number of matches: 2

1 125-127 TER

2 284-286 SFR

[3] PDOC00006 PS00006 CK2\_PHOSPHO\_SITE

Casein kinase II phosphorylation site

Number of matches: 2

1 161-164 SILE

2 173-176 SDGD

[4] PDOC00008 PS00008 MYRISTYL

N-myristoylation site

Number of matches: 6

1 9-14 GTESTT

2 17-22 GNDQAL

3 44-49 GLVGNG

4 200-205 GSSLAL

5 215-220 GLPLTR

6 236-241 GLPFGI

[5] PDOC00029 PS00029 LEUCINE\_ZIPPER

Leucine zipper pattern

Number of matches: 3

1 24-45 LLCGETLIPVFLILFIALVGL

2 31-52 LIPVFLILFIALVGLVGNFVL

3 216-237 LPLTRLYLTLILTIVLVFLLCGL

[6] PDOC00210 PS00237 G\_PROTEIN\_RECEPTOR

G-protein coupled receptors signature

115-131 AGLSMLSTVSTERCLSV

## Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	30	50	2.256	Certain
2	63	83	1.387	Certain
3	101	121	1.904	Certain
4	144	164	2.070	Certain
5	185	205	2.359	Certain
6	221	241	2.328	Certain
7	265	285	0.965	Putative

3/8

## FIGURE 2, sheet 2 of 2

## BLAST Hits and Alignment to Top Hit:

Sequences producing significant alignments:

	(bits)	Value
gi 134079 sp P23749 RTA_RAT PROBABLE G PROTEIN-COUPLED RECEPTOR...	125	7e-28
gi 6981186 ref NP_036889.1   MAS1 oncogene >gi 135921 sp P12526...	111	1e-23
gi 4505105 ref NP_002368.1   MAS1 oncogene >gi 135920 sp P04201...	109	6e-23
gi 6678804 ref NP_032578.1   MAS1 oncogene >gi 266505 sp P30554...	106	5e-22
gi 2118485 pir  S51001 transforming protein mas - mouse	104	1e-21
gi 547920 sp P35410 MRG_HUMAN MAS-RELATED G PROTEIN-COUPLED REC...	104	1e-21
gi 3642615 gb AAC36503.1  (U86378) anaphylatoxin C3a receptor [...	63	6e-09
gi 6831498 sp O88680 C3AR_CAVPO C3A ANAPHYLATOXIN CHEMOTACTIC R...	63	6e-09
gi 4504093 ref NP_001497.1   G protein-coupled receptor 32 >gi ...	61	2e-08
gi 3023641 sp P97468 CML1_MOUSE CHEMOKINE RECEPTOR-LIKE 1 (G-PR...	59	8e-08
gi 3023791 sp P79189 FMLR_MACMU N-FORMYL PEPTIDE RECEPTOR >gi 1...	59	8e-08

>gi|134079|sp|P23749|RTA\_RAT PROBABLE G PROTEIN-COUPLED RECEPTOR RTA  
 >gi|112279|pir||A35639 G protein-coupled receptor RTA -  
 rat >gi|206810|gb|AAA42087.1| (M35297) G-protein coupled  
 receptor [Rattus norvegicus] Length = 343

Score = 125 bits (311), Expect = 7e-28

Identities = 88/275 (32%), Positives = 126/275 (45%), Gaps = 12/275 (4%)

Frame = +3

Query: 222 FLILFIALVGLVNGFVLWLLGFRMRRNAFVSIVLSLAGADFLFLCFQIINCLVYLSNEFF 401  
 ++ L + L GLVNG VLV GF ++R FS+Y L LA AD ++L + + L+ + F  
 Sbjct: 49 YIFLLCLCLVGNGLVLWFFGFSIKRTPFSIYFLHLASADGIYLFSAVIALLNMGTF 108

Query: 402 CSISINFPSTFTTVMTCAYLAGLSMLSTVSTERCLSVLWPIWYRCRRRPRHXXXXXXXXXX 581  
 S V C + AG+S+L +S ERC+SV++P+WY RRP+  
 Sbjct: 109 GSFFDYVRRVSRIVGLCTFFAGVSLPAISIERCVSVIFPMWYRRRPRKRLSAGVCALLW 168

Query: 582 XXXXXXXXXXXXEGKFCGFLFSDGSGWCQTFDFITAAWLIFLF---MVLGSSSLALLVRLC 752  
 FC FL + C D L FLF MVL LAL++ + C  
 Sbjct: 169 LLSFLVTSIHNYFCMFLGHEASGTACLNMDISLGILLFLFCPLMVL--PCLALILHVEC 226

Query: 753 GSRGXXXXXXXXXXXXXXXXXXXXXPF-GIQWFLILWIWK-----DSDVLFCHHP 905  
 +R + GI WFL W+++ +D+ C  
 Sbjct: 227 RARRQRSAKLNHVLAIVSVFLVSSIYLGIDWFL-FWVFQIPAPFEYVTDLCIC---- 281

Query: 906 XXXXXXXXXXXXANPIIYFFVGSFRKQWRLQQPILKLALQALQDIAE 1046 (residues 35-309 of SEQ  
 ID NO:2)

A PI+YF G + Q RL +P L++ QRAL+D AE  
 Sbjct: 282 -----INSSAKPIVYFLAGRDKSQ-RLWEP-LRVVFQALRDGAE 319 (SEQ ID NO:4)

## Hmmer search results (Pfam):

Model	Description	Score	E-value	N
Scores for sequence family classification (score includes all domains):				
Model	Description	Score	E-value	N
CE00141	CE00141 Mas_proto-oncogene	121.6	2.4e-38	1
PF00001	7 transmembrane receptor (rhodopsin family)	57.5	2.1e-17	1
CE00200	CE00200 DEZ	21.1	3.5e-05	2
CE00175	CE00175 Purinoceptors	20.2	7.6e-05	3

## FIGURE 3, sheet 1 of 5

```

1  AAAAATCCTC AATAAAGTAC TGTCAAACCG AATCCAGCAG CACATCCAAA
51  AGTTTATCCA CCATGATCAA GTCGGCTTCA TCCCTGGGAT GCAAGTCTGG
101 CTCAAAATAC GCAAATCAAT AAATGTAATC CATCACATAA GCAGAACCAG
151 TGACAAAAAC CACATGATTA TCTCAATAGA TGCAGAAAAG GCTGTCGATA
201 AAATTCAACA CCCCTTCATG CTAAAAACTC TCAATAAACT GGGTACTGAT
251 GGAACATATC TCAAAATAAT AAGAGCTATT TTTTATAAAC CCACAGCAAA
301 TATCACACTG AATGGGCAAA AGCTGGAAGA ATTCCCTTTG AAAATTGCCA
351 CAAGACAAGG ATGCCCTCTC TCACCAC'TCC TATTCAACAT AGTATTGGAA
401 GTTCTGCACA GAGAAATCAG ACAAGAAAAA GAAATAAAGG GTATTCAATT
451 AGGAAAAGAG GACGTCAAAT TGTCCCTGTT TGCAGATGAC ATGATTGTAT
501 ATTTAGAAAA CTCCATTGTC TCAGTCCCAA ATCTCCTTAA GCTGATAAGC
551 AACTTCAGCA AAGTCTCAGG ATACAAAATC AATGTGCAAA AATCACAAGC
601 ATTCCCATAC ACCAATAATA GACAAACAGA GAGCCAAATC ATGAGTGAAC
651 TCTCATTTAC AATTGCTACA AAGAGAAATA AATACCTAGG AATACAATT
701 ACAAAGGATG TGAAGGACCT CTTCAAGGAG AACTACAAAC CACTGCTCAA
751 GGAAATAAGA GAGGATACAA ACAAAATGGAA AAAAATTCCA TGCTCATGGA
801 TAGGAAGAAT CAATGTCGTG AAAATGGCCA TACTGCCCAA AGTAATTTAT
851 AGTTTTAATG CTATCCAGT CAGGCTACCA TTGACTTTCT TCACAGAAGT
901 AGAAAAAACT ACTTTAATT TCATATGGAA ACAAAAGAGA TCCTGTATAG
951 CCAGGACAAT CCTAAGCAAA AAGAATAAAG CTGGAGGCAA CACACTACCT
1001 GACTTCAAAC TATACTACAA GGCTACAGTA ACCAAAACAA CATGGTGCTG
1051 GTACTAAAAA ATATATATCG ACCAGTGGAA CAGAACAGAG GCCTCAGAAA
1101 TTATGCCACA CATGTACAAC CATCTGATCT TTGACAAACC TGACAAAAAC
1151 AAGCAATGGG GAAAGGATTC CTTATTTAAT AAATGGTGTT GGGAAAACTG
1201 GTTAGCCATA TGCAGAAAAC TGAAACTGGA CCCCTTCCTT ACACCTTATA
1251 CAAAAATTAA CTCAAAATGG ATTAAAGACT TAAACATAAG ACCTAAAACC
1301 ATAAAAACCT AGAAGAAAAC CTAGACAGTA CCATTCAAGG CATAGGCATG
1351 GGCAAAGACC TCATGAATAA AACACCAAAA GCAATGGCAA CAAAAGCCAA
1401 AATTGACAAA TGGGATCTAA TTAAACTAAA GAGCTTCTGC ACAGCAAAAG
1451 AAATATCAT CAGAGTGAAC AGGCAACCTA CAGAATGAGA GAAAATTTTT
1501 GCAATCTATC CATCTGACAA AGGGCTAATA TCCAGAATTT ACAAGAAACT
1551 TAAACAAATT TATAAGAAAA AATAAACAACT TCCATCAAAA AGTGGGTGAA
1601 GAATATGAAC AGAGACTTCT CAAAAGAAGA TATTTATATG GCCAACCAAC
1651 ATGAAAAAAT GCTCATCATC ACTGGTCATT AGAGGAATGC AAATTAAAAC
1701 CACAATGAGA TACCATCTTA CACCAGTTAG AATGGAAATC ATTTAAAAGT
1751 CAGGAAACAA CAGATGCTGG AGAGGATGTG GAGAAATAGG AACGCTTTTA
1801 CACGGTTGGT GGGAGTGCAA ATTAGTTCAA CCATTGTGGA AGACAGTGTG
1851 GCGATTCTC AAGGATCTAG AACCAGAAAT ACCATTTGTC TCAGCAATCC
1901 CACTACTGGG TATATACCCA AGAGATTATA AATCATTCTA CTATAAAGAC
1951 ACATGCACAC ATACGTTTAT TGCAGCACTA TTAACAATAG CAAAGACCTG
2001 GAACCAACCC AAATGCCCAT CAATGATAGA CTGGATAAAG AAAATGTGGT
2051 ACATATACAC CATGGAATTC TATGCAGCCG TAAGAAAAGG ATGAGTTTAT
2101 GTTCTTTGCA GGGACGTGGA TGAAGCTGGA AACCATCAT CTGAGCAAAAC
2151 TAACACAGGA ACAGAAAACC AAACACTGCA TGTTCCTACT CATATGTGGG
2201 AGTTGAACAA TGAGAACACA TGGATACAGG GAGGGGAACA TCACACACTG
2251 GGGCCTCTCA GGGGGTGGGG AGCTAGGAGA GGGATAGCAC TAGGAGAAAT
2301 ACCTAATGTA GATGACGGGT TGATGGATGC AGCAAACCAC CACGGCATGT
2351 GTATACGTAG GTAACAAACC TGCACGTTCT GCCCATGTAT CGCATAACTT
2401 AAAGTATAAT TTTAAAAAAA TAAAATAGAA CTTCTAGAAA TTAGTAATAT
2451 AAAAAATTGA TTAAACTTC AGTGGATAGA TTAAACAGAT GATTAGACAC
2501 AATGGAAAGA ATCTAGTTAA CTGGAAGATA CATCTGAAGA AATTACCCAG
2551 AAAGCAGCCT CAGAGACAGA GACGAGACAT AAGGAAAAGG TTAAAGAAAT
2601 AAGGATGATA GAAGAAGAGG GCTAACGTAT GTTGAATTAG ACTTCCAGAA
2651 AGAGATAATA CTAAAAGTGA TAATGAGAGC TAATAATTTT ATAGTACTTA
2701 CTATGTGCTC CTATTGTAAA TCATTTATAT ACATTAAATT ATTAATCTTC
2751 ACAGCAAAAC TATAGATTAA TAATGTTATC CCCATTTTAC TGATGAGGAA
2801 ACAAACATAA CAGAGTTACA AAATCTGGCT AAGGTCCCAT AGGAGAAATG
2851 CCAATAGGCC TGCTCTTGTG TTCTTCACCG TTATGTTAAT ATTTAAAGAG
2901 ATTATAGCAT TAAATTGTTT AAAAATGTTA AATGATGATA ATTTTAAAT
2951 CTAGGAAACT CCACAAATCC CAAACATGAT TTGAAAAAGA GTCACCAGAT
3001 ACAATTTGGT GAGAATGTAT CATAGCAGAG AAAAATAAAA GCTCTTAGAG
3051 GAAAAAAGG CAGATCACCA CAAGCAAATG GCAACCTGAC TGATGGCCAC
3101 AACACTGGCA GTCAGAACAT GGATCTTCAA TGTGCTGGAA GAAAGAAACT

```

5/8

## FIGURE 3, sheet 2 of 5

```

3151 ATCACCCCTAC AATGCCATGC CTGGA AAAAC CATCTTTAGA AGACGAGGGT
3201 GAAACAATGA CATTTTCAGA TAGATAAAAG GAGGCATTTA CCATGCAGAG
3251 ACTCTCACAG AAGGAACTCC TGAAGAATGT GTTTTGAGAA ATGGAAGATA
3301 TGAGATGTAA AATGGGAGGA TGAGCAAAAG CATTGGTGAT TATGTGGATA
3351 AATCTGAATA AACACTGACT ATTTAAACA ATAATAAGTG TTTCATTTGT
3401 TGGAGGGAAA CAGAATCAAA ATCAAATAAT AGAGCACAGT AGGTTAGACA
3451 GTAACGGAAG TTAAAGCTTT CTAAGGTCTT GCTGCTGCTC ATGAGGAAAAG
3501 CAGATTTCTG GGGACAAATG CATGGTTTAC CCACTGCACC AAAGTATAAT
3551 ACAAGAAAAT GCTGAGTCTC AGAACTCGCC TCAATCAACA CATGTTCCCTT
3601 GTCTGAGGTC ACCCAGCTGC TCTGGAATGC AAAGGGGCAT GATCTTGATAT
3651 GGTGTTTCATC ATTGTGTCTT CATTGCCTGG AACGGTACCT GACATATCAT
3701 AGGCAAAACA TAAACAGTGT TGAATGAATG AATAAATGAG TGAATAATTT
3751 CTTCTGATTG TCTTAAATGC TTCATAATTG TTACTTATAA CTCAATTTGG
3801 CTGCAGCATA GGCACATGAA AGGAAGTCAT AGGAGATAAG GCTGAAAATA
3851 TTGGTTGGCT CTAGATGATC TAAAAATAAT CTAGCTAATA AATATCATGA
3901 TGTAAGAGGA GAAACGTGAC AGATGGTGCA TGTGATGTGC TAAGTCCCCT
3951 ATATTTTATA AGGAGACTCA ATATATCTAG TTGTTAACAT TGATAAATCA
4001 AAAATAGAAA TGTTGTTTAA TTACGCAAAG CTTTTTTAAA AAAAAGAATG
4051 GAAACTGTAT GCAGTGGTTG CTTTTGGGGA ATGAGACTTC GAGAGACTTT
4101 GCATTTTCATT TTATACACCT TAGTGTCATT TGAATAAAC CTGATTTTGA
4151 AAGGTCAAGT AAGCCAGGGC CTGGAATGCC AAGCAGAAGA GGCAGACTA
4201 TTCTTTAGCT ATGAAAGCTC ACGTCGCTTG ATTGGTCTGC TTCCTGGATG
4251 CTCATTTGGG GCAATGAAAG ACTTCAGGAC GGTGATTCCCT TGCAGGGAGT
4301 GGGGCCGGAG AGGGTAGGGG AGAAACCAAG ATAAGTGATT GGAGTCAGGG
4351 CAAATGCTGA GAAACATCTC ATGTGGTTTC CTGTGTCATA CTCTATTTT
4401 CCAGCACCCCT CTGTACAGGA AATCACAGGG TTGTTGAGCA TCTTGATAAG
4451 CTTCACTCAG TCATCAGGTT TCACTCCAGA TCCTGCAGGC ATCTCCCAT
4501 CCTCAGCTGT TTGCCAGTCC CAGGAAAGCA CTTCTCAACT CACCAACTCC
4551 AGTAGAAAGA AGGGTGTTAA GGTAAGAGTT TGTTCAGAA CCATCTTCTT
4601 TCAAAGGCAG TTTTGGTACC TTAGTCCACA TCCTGATGAA TTCATTAGGA
4651 AAAGAAGACA GGAAGCTCCT TTGAGGAAGC TTTGGATGCT GGCAGCTTAG
4701 ATGGTGGGAG ATCCGAATGG GGTCAGCAG GTTCGTCCCA CTCTCTTCT
4751 ATCTACCAAT ATTCTCCCTT TCTCTCGAC CATTCAAATC TGAAACTATG
4801 GAGAGAAAGA TTTTGCTGAT TTTAATTTAG GGCAGGGCTA GAATGAAGTC
4851 ATTACTTCCC TAAGCTAAAT TGTATTTTAA AACCTCAGGT TGCAGTGATT
4901 TTCTTGTTAT TCAAAAGGTT TTCTCTCAT ATCTGTCATT GTCCAGCAG
4951 AAAACACACA TCACACACTC TAAAGGGGTA GTTTAGAGGA GTTTAGTGAA
5001 GAGACTGTTT ACAACGGCGT GAACAGGGAT TAAGGAATGA TGAGCCCTC
5051 AAAGGCAAGT GATAGTGGAG GGCTGTTACC GTTGCTAGGT CTGAAGTGAT
5101 AAGGAGAAG AGAAGTTCTG AGAACTCAGA ATGTGCTGTA GCGTAACTG
5151 TAACTGTAAC TATAGCTGTG GCTACAGGAA AGGGCCACCA GACAGGGCTA
5201 TGTCCTTAGG TAGAAAAACA CTGCCACTGC CAACTCACAG CCCTTCAGGG
5251 CGCAGGGAGA GAGCCAGGAA ATTTTTTAAA AATCATCTC CAATCTCCTG
5301 TCAATGTGTC CCTTTGGCTG AAAAAAAAAG TCACCCTCCA ATCTCCTGTC
5351 AATGTGTACC CTTTGGAGCC TGAGTGAAAG ACAGCCCAT TACGAGGCAC
5401 AGACATGTCT CCTCCAGGA TGCAAAGGCA GGTAGAGAAG GATGGACATG
5451 AGATCCTAAT AGCAAATAGG AAAAGTCCAT TTCTTCCAT AGCCTCTTCT
5501 CAGTTTGCTT TTCCTGAGAC TTTCTCTATT AATGTGATTG AATCAATTTT
5551 TCATTCTATC ACCTCCCTTT TTTTAAATTG TTTGACCTC CCATCTAGTG
5601 TCACTTCTTT GGACTAGTCT CTCACATCA TCATAAATGC CTTGAGAATG
5651 GAATGTGGTT GGGAAAAAAA GGGATTGGGA GTACATAGGT ACTCCAGCT
5701 ATAAGTACAC AGGTATGTGT AATGTAAGTT TATGATTTTG GCTTCTCTAA
5751 AAAGAAACCT ATTATAGTGA TAATCAAAAA GAATGTTTAT TAGACTTGGG
5801 CTGATAACTC TAAGACATCA AACATCAAGG CTCCTTCCAA TTTCTATTTT
5851 CCCATCCCTA AAGCGGTACC CTTACCCTCA TGGTCCTACA TGGCTCCCAA
5901 ACACATCAGC ATTCCAGCCA GAGAGAAGAG AAAAAAAGAA AGAGTACAGG
5951 AGGAGCTGGT TATATCACTT TCTCTCACCT CCCCTTGACC ATCACATATG
6001 GTTACATGGT GCCATTAGCA TAAAGGGAAG CTGAGAAATG GGCCTTTTTT
6051 TGATCTAAGA CTCAGAATTA TTATGTGTGA GAAAAGGAGA ATGGCTTTTG
6101 GGGCAACAAG CAGTCTCTAT AGCCCGATTG TCTTTTAGCA TCTGTGCTCA
6151 CCTGGAGATT TTTCTCAACA CTGCCTCACA AATAGTGAAT GATGTAGACA
6201 TGGAACAGAA ATAAACCTAG GACAGGGCCC CAGGAGACTG GAGCTGGTAT
6251 TGGACCTGCT TTTCACCATG TGATTCAGGA AACTCTTTTG TACATGTTAA

```

## FIGURE 3, sheet 3 of 5

```

6301 GCCTGTCTTT CTTTGTAA TGAAGCTATT ACAATAGATG GCCTCCAAGG
6351 GATATGTGTT GGAGGCTGGG GATATAATAG TGAACAACCA CAACAAAAAT
6401 GTTGAGCTTA ATGAGGGAAG TAAAAGAAAA AACAGACATA CACTTAAAAA
6451 TACAGTGCA GGGCCGGGCA CGGTGGCTCA CGCTGTAAT CTTGGCACTT
6501 TGGGAGGCCG AGGCGGGCAG ATCACAAGGT CAGAAGTTCG AGACCATCCT
6551 GGCTAACACA GTGAAACCCC GTCTCTACTG AAAAAAATAA TACAAAAAAT
6601 TAGCCAGTCG TGGTGGTGGG CACCTGTAGT CCCAGCTACT CGGGAGGCTG
6651 AGGCAGGAGA ATGGCGTGAG CCCAGGAGGC GGAGCTTGCA GTGAGCCGAG
6701 ATCGAGCCAC TGCATCCAG CCTGGGTGAC AGAGCGAGAC TCCATCTCAA
6751 AAAAAAATAA AAAATACAGG TGCAAACTGT GTTCAGTGCC ACGAATAAAA
6801 AGGCCAACCT GCCCTGAGGG AGGACAACAG AAGTGGTTTT CTTCATATC
6851 CTCACAAGGC CAACCAGGGC TACAGGAACT AAGACAGCAT GTGATGGATG
6901 CATTCACTCC TTGATGCCAA TTCTGCACCC CCATGGAGGA CACATTCAT
6951 TCCCCCTGG AGGCCTCCAA CCCATGAGAG TTTTCTGCAG TGTGGAGTGA
7001 TTGACGGCTG TTTTCTGAA GCAAGAAAGA GCTCTGTCAT TCAGTTGCTT
7051 ATTCTCAGTT GGCTTCTATT CATATCTCTT CTCCTTATTT GATTAATCTC
7101 TCTTCCTTAC TGTTTCTTAA AGTTGCTTAA TCTAGGCTTA AAAAACTGCA
7151 AATTCGACAT TCCATTTTGG ATTTGATATG AGCATTACTT GTTGCTTTTC
7201 TCCTTTCTTC ATGTGTAGTT GGTATTGAAC CCTTGCAATG GGGCTGCAAG
7251 GCTGGTTAAG AACAAGCCCT CTGACACAGA CCTGTGGGTT CAAATGCAGG
7301 CTCTACCACT TGTTTGTGAG ACCTGAAGCA TATTACTCAA TACATTTGTG
7351 CTTTAATCTC CTCTCTGTA AAATGGGACT AATAATAGTA CTGACCTCAC
7401 CAGGCTCTGA AGAAGATTGA AAGAGGCTAT ACCTGAGGAG CACTCAGCCA
7451 GCTGCTATTA TTCACCCCAT TATGAGTAAG GGCTCCTTCC CAAGTGTAAT
7501 TTGCATAAGC ATCTGCCTTT TCCCCCATAA GCCCTGCACA GTGTAGACAT
7551 GGTTCCTTTT TCTAATTTCC CACTCCTTTC CCAAATAAT CCTGTTTACA
7601 AATGCTTTAA TTGCAAGA AAGCTCAAAA TCATCTGGGG GAAGTAATGG
7651 GTCCACCTTA GATGTATTCA TACCACAACA CCCTCTAGTT ACCTCTAGAC
7701 AATAGACACT CTCAATTGCT GCCTCTCTCT CTTTTTCCCA AATTGCCCCC
7751 TTTTCTCTCA AGATGCAGAG AAATTCAGAT AAACCTGAA TCCTCTCCAT
7801 AAATCGAAG TTTAACTCT TTTGGGTGCT GAGCAAAAGC CATATACTAG
7851 CCCTTCTGG ATTGAAGATA TGACAGTGGC CATTATTAC TATCTGTGTA
7901 GGCAGAAAGA ATCCCTTCAA ATACACCAGA GCTTTCCAG AAGGAAACAT
7951 GAAGAATGAT GTTTTCTTCA TTTCTTTACA TGTGTGTTTT TTAGTGACTA
8001 ATTTCAATTT AAAATGTTGG TGTTCACCTT AGCCCAATTT TTCAACTACT
8051 CTTGGCTCCC CATGCATGCA TGCATGCATT CATTATTCA TTTATATGTG
8101 GTACAAACTT CTTGACCCTC CACTATGCCC TCACATTGAG CTAGGCACTG
8151 AGGACACAGA AATGGATAAG AGAAAACTC TTTCTGTTTT CATAAAGCTG
8201 ACCACTGAG CAATTTCTAG CCTGACAGGC CTGGCTGTCC TCAGACCTGT
8251 CCTAGCAAGA AGTCACATCT CCGTTATAAC CCTAAATCAT GATCTCCATT
8301 CTGACTGTTG ATGGACCCCT CTGAGTCATT AAGGGTGGCC ATGGGTCATG
8351 GCCTGGGTCA TTTCTTCTGG CCCCTGAGAG GAAATCTGTA TGCCAGGATA
8401 AAAAGAAATG TGAACCTAGA GTCTGTCTC TTGCTGCCCC ACATATTCTG
8451 CCTAGACTTT TCTCTTCTG CCTCCAGGAG ACATGACACA GTGACAGTGA
8501 GTGGGGGTGT TTGGGCTTA GAATATTCCC ATACCAGCAG AGGATCTACC
8551 TTGTAATAGA AAGAGCCCAG GATTAGAGT CAGCAAGAAA TGAGTTTGAA
8601 TCCAGGTGCT AGGACTCCCT GGCCCTTAAT AAATGACTTA ATCTCTTCAA
8651 GCCTCTGATT TCCTCTCTG TAAAACAGGG GTGGTAATTA CCACATAACA
8701 GGCTGCTCAT GAGAATCAGT GAACATGCAG CAGGTGCTCA AGTCTTGTTT
8751 TTGTTTCCAG GGGCACCAGT GGAGGTTTTT TGAGCATGGA TCCAACCACC
8801 CCGGCCTGGG GAACAGAAAG TACAACAGTG AATGGAAATG ACCAAGCCCT
8851 TCTTCTGCTT TGTGGCAAGG AGACCCTGAT CCCGGTCTTC CTGATCCTTT
8901 TCATTGCCCT GGTCCGGCTG GTAGGAAACG GGTGTTGCTC CTGGCTCCTG
8951 GGCTTCCGCA TGCGCAGGAG CGCCTTCTCT GTCTACGTC CTAGCCTGGC
9001 CGGGGCCGAC TTCCTCTTCC TCTGCTTCCA GATTATAAAT TGCTGGTGT
9051 ACCTCAGTAA CTCTTCTGT TCCATCTCCA TCAATTTCCC TAGCTTCTTC
9101 ACCACTGTGA TGACCTGTGC CTACCTTGCA GGCTGAGCA TGCTGAGCAC
9151 CGTCAGCACC GAGCGCTGCC TGTCCGTCT GTGGCCATC TGGTATCGCT
9201 GCCGCCGCC CAGACACCTG TCAGCGGTG TGTGTGCTC GCTCTGGGCC
9251 CTGTCCCTAC TGCTGAGCAT CTTGGAAGGG AAGTTCTGTG GCTCTTATT
9301 TAGTGATGGT GACTCTGGTT GGTGTCAGAC ATTTGATTT ATCACTGCAG
9351 CGTGGCTGAT TTTTTTATTC ATGGTCTCT GTGGGTCCAG TCTGGCCCTG
9401 CTGGTCAGGA TCCTCTGTGG CTCCAGGGGT CTGCCACTGA CCAGGCTGTA

```

## FIGURE 3, sheet 4 of 5

```

9451 CCTGACCATC CTGCTCACAG TGCTGGTGTG CCTCCTCTGC GGCCTGCCCT
9501 TTGGCATICA GTGGTTCCCTA ATATTATGGA TCTGGAAGGA TTCTGATGTC
9551 TTATTTTGTC ATATTCATCC AGTTTCAGTT GTCCTGTCAT CTCTTAACAG
9601 CAGTGCCAAC CCCATCATTT ACTTCTTCGT GGGCTCTTTT AGGAAGCAGT
9651 GCGGGCTGCA GCAGCCGATC CTCAAGCTGG CTCTCCAGAG GGCTCTGCAG
9701 GACATTGCTG AGGTGGATCA CAGTGAAGGA TGCTTCCGTC AGGGCACCCC
9751 GGAGATGTCG AGAAGCAGTC TGGTGTAGAG ATGGACAGCC TCTACTTCCA
9801 TCAGATATAT GTGGCTTTGA GAGGCAACTT TGCCCCTGTC TGTCTGATTT
9851 GCTGAACCTT CTCAGTCCTG ATTTTAAAC AGTTAAGAGA GTCCTTGTGA
9901 GGATTAAGTG AGACAGTGCC TATGAAACAA AACTAAGTG CAGTGTCTCT
9951 GGAACCTCCT TACTCACAGG CTTCCACCAC AGCCTATGA GAGCTTTGCC
10001 AACTCTGCGG TCCATGACTG TTCCCACTTT TAATGAATCC TACCTTTCGC
10051 AGAAGGCTGA AAGCAGGCA GAAAAGTCT ACATTTCTTT GGTCAGTCA
10101 CTTGATAGGG ACTCAAACAA TGTATATTT TTAATTAATT TCTTTTCTC
10151 TTCCGTACAA TTTCTGTCTC AACAAAATTA GAAGAATTAA ATTTAAACT
10201 AGCTCCAAA GAGCAGTCGT CTTTCATTTT GGCAGACCTT AGAATATCCC
10251 CCTAGCTTAA TAAATCTTTG TTGAATGGCT TAATGAATGA ATAACTGGT
10301 TAACGTTTAA GTTAAACCTC TGAAAAGTCT CCATTTACCA GATTTGAGTC
10351 ACTAAATTTA TTGCTTTCAC TACTTTTGAA TTTTGCAAAC ATGAAATTAA
10401 GTTTTATAAT TAGATAAATC AATGTCAACA CATATTTAAA GTTTGAGGTA
10451 CACTGTCTTC CTGTGGTTTC CTTTCACATG CCATCCCTTA AAATCCCAGC
10501 TACACGCCTT CCCATTCTTT CCCCTTTGCC TTTGTTCTAA TCTTCCCTCT
10551 CTGGGGGCTC TCTAATTCGT CCTGGAAGTT TCCAATGGTC TTATAGACTC
10601 CATGTTCTTG GAGGACAGGC TGTATGTCAG ATTTACCTTT TATTCCGAAG
10651 AACTCGGAGC ATTTATTTTG TTAATTAAAT TGCACATATT TTTAAAAGTT
10701 ACGTGTTCCT CAGAATAAAA TACTAATTGT AAATGCTGCA TCTTTAATA
10751 ATTTTTTATT ATTTTAAATT AAGGTATAAT TAAGAGAATG ATATGCACAG
10801 ATCATAAATG TACAGCTTAA TGAGTTTGT TAAATGAATG TGAATCAAAC
10851 TTCATTCAAG ACAAAATATT TTCATCATCT CTAAAAGTTC CCCTGTGCTT
10901 CCTCATCTCT CTGCCCCCA GAGGCAACCA TTGTTTCTAT TTCAATCAAC
10951 AAATATTAGC TTGCCTGAC CTTTAACTGT ATCTTTTATG AATCATAATA
11001 TGTGTATTTT TTGTTTCTAG CTCTTGCACT CAACATAATA TTTTGA (SEQ ID NO:3)

```

## FEATURES:

Start: 8786

Exon: 8786-9778

Stop: 9776

## SNP information for allelic variation:

Position	Allele 1	Allele 2	Context
8692	c	t	Aaacagggg[c/t]ggtaattaccacataaacaggctggatcatgaaatcagtgaacatg cagca (SEQ ID NO:5)
8716	a	g	aaacaggggacggttaattaccacataaacaggctggatcatga[a/g]aatcagtgaacatg cagcaggtgctcaagtcttgtttttgtttccagggg (SEQ ID NO:6)
8970	a	g	ggtaggaaaacgggtttgtgctctggctcctgggcttccgcatgacgagga[a/g]cgcc ttctctgtctacgtcctcagcctggccggggccgacttctcttcc (SEQ ID NO:7)

## FIGURE 3, sheet 5 of 5

				Protein		
POSITION	Allele 1	Allele 2		Position		
8692	c	t	Intron			
8716	a	g	Intron			
8970	a	g	Exon	62	S	N

## Map Position:

#	SHGCNAME	CHROM#	LOD_SCORE	DIST. (cRs)
1	SHGC-34534	11	11.64	10
2	SHGC-5996	11	11.64	10
3	SHGC-548	11	11.64	10
4	SHGC-32644	11	10.69	15
5	SHGC-11901	11	10.69	15